Supriya Tumkur Suresh Kumar

GitHub ***LinkedIn*** ***Website***
supriyats28@gmail.com | 585-319-9510

TECHNICAL SKILLS

Programming – Python, SQL
Machine Learning – Regression (Linear, Multiple-Linear, Polynomial, Logistic, SVR, Random Forest); Classification (K-NN, SVM, Random Forest, Naïve Bayes); Clustering (K-Means, DBSCAN, Hierarchical); Deep Learning (NLP, RNN, LSTM, CNN, GAN); Gradient Boosting.
Database & Cloud – MSSQL, PostgreSQL, Google Cloud Storage (GCS), BigQuery, Netezza, Google Cloud Platform (GCP).
Tools & libraries– Pandas, NumPy, SciPy, scikit-learn, TensorFlow, PyTorch, Beautiful Soup, Selenium.
Data Integration, Visualization & Deployment – Alteryx, Tableau, Seaborn, Matplotlib, Flask, Heroku.
Research & Analysis – Exploratory Data Analysis (EDA), A/B Testing, Hypothesis Testing, Data Mining.

EXPERIENCE (5 years)

PHOENIX program, Wayne State University, Detroit MI – *Data Scientist*

The Population Health Outcomes and Information Exchange (PHOENIX) Program
June 2020 – March 2021

- Deployed a text similarity Natural Language Processing (NLP) technique on GCP to create a unified covid-19 survey dataset by combining 20 disparate surveys, reducing 15 hours of manual work per week.
- Built web scraper to automate weekly and monthly data ingestion using BeautifulSoup, Selenium, Google Cloud Functions and Scheduler, saving external vendor costs of $30k annually.
- Utilized Google Compute Engine to schedule python scripts for cleaning and pre-processing large datasets, resulting in improved data pre-processing speed by 80%.

Trustworthy AI Lab, Wayne State University, Detroit MI - *Machine Learning Research Assistant*

September 2019 – March 2021
Mentor Dr. Dongxiao Zhu

- Collaborated to design new training scheme for model compression ensuring adversarial robustness, explainability and personalization for NLP applications. **preprint arXiv:2101.05624**.
- Implemented On-Device model personalization-trained global model on a sentiment dataset collected from 5000 users and then fine-tuned it on each user's private training data.
- Extracted static geospatial features from 20,000 geocoded POIs in Detroit using PCA and clustering and explained it in terms of POI types to learn semantic similarities between regions for efficient public health surveillance. ***(Clients- School of Medicine Wayne State University & Detroit Department of Health)***
- Curated Yelp review dataset of 50k restaurant reviews and ratings using Python and Beautiful Soup for sentiment classification.

Neo Prism Solutions LLC, Schaumburg, IL - *Data Analyst Consultant*

June 2015 – June 2018
Client- Sally Beauty Holdings

- Built an end-to-end analytics pipeline for data extraction, processing and reporting of payroll and census data using SQL and Alteryx, saving the company $500k annually in reduced overtime cost.
- Developed an interactive dashboard to report transactions and sales of ~3000 stores in real time and

communicated the key findings to the leadership.

Client- Adams Street Partners

- Trained 20 executives to build ad-hoc Tableau reports reducing 10 hours per week of manual work to generate single use reports.
- Developed a client interactions dashboard with details of 30 days prior and subsequent interactions, resulting in 10% increase in deal completion.

Client- Blue Cross Blue Shield of Illinois

- Created tableau custom geo code map based on three-digit US Zip codes for tracking plan providers' participation.
- Built workflow for design, development and maintenance of ongoing metrics, reports, and dashboards to drive key conclusions on each Blue Distinction Specialty Care.

Client- Compassion International

- Developed Tableau dashboard to track high-risk hourly transaction errors logged by Enterprise Service Bus (ESB), reducing the error handling time by 60%.
- Created reusable SQL queries to port daily ESB audit data into Netezza resulting in improved production database performance by 12%.
- Collaborated with marketing team to analyse data using Python and SQL to generate Tableau visualizations for latest sponsorships and campaign; led to $50k in new sponsorships.

DATA SCIENCE PROJECTS

- **Adversarially robust sentiment analysis** *(GitHub)*
  - Used Adversarial Training approach to build a sentiment classification model that is robust against adversarial attacks.
  - The trained Bi-directional LSTM model achieves 84.7% accuracy on generated adversarial samples.
  - Application: The robust model can be used by businesses that rely on public reviews to provide product recommendations to prevent attacks on their models and provide reliable service to its customers.
- **Liver failure prediction using determinants of health** *(GitHub)*
  - Trained an XGBoost model to predict liver failure using demographic and personal health variables.
  - Handled the dataset imbalance with oversampling.
  - Application: Insurance providers can use this approach to predict health behaviours of their members and connect them to social and environmental services to reduce overall medical costs.
- **Semi-supervised approach to detect distracted driver using driver dashboard images** *(GitHub)*
  - Implemented a Semi-Supervised GAN to build a multi-class classification model using both labelled and unlabelled 2D images to distinguish between 10 distracted driving activities.
  - Trained the model with 15k labelled and 40k unlabelled data and achieved evaluation accuracy of 94.7%.
  - Application: The trained model can be used by OEMs to improve auto safety and semi-supervised image classification reduces the cost of collecting expensive labelled data.

EDUCATION

Master of Science in Computer Science — Current
Wayne State University, Detroit, MI — GPA - 3.89

Master of Science in Electrical and Computer Engineering — May 2015
Oklahoma State University, Stillwater, OK — GPA - 3.36

Bachelor of Engineering in Telecommunication Engineering — June 2012
Siddaganga Institute of Technology, Karnataka, India — GPA - 3.77